图书情报工作
第66卷 第12期 2022年6月

面向考古类型学的出土陶器器形的知识表示与语义关联构建*

■ 韩牧哲[1] 高劲松[1] 李钰[2,3]
[1] 华中师范大学信息管理学院 武汉 430079 [2] 四川大学考古文博学院 成都 610225
[3] 四川文化艺术学院美术设计学院 绵阳 621000

摘 要：[目的/意义] 面向考古类型学，提出一种适用于出土文物特征描述的知识表示模型，并在此基础上提出相应的语义映射和本体扩展方法，以突破传统类型学方法造成的语义壁垒，促进知识的共享与重用。[方法/过程] 首先，对考古类型学思想及其传统方法造成的语义壁垒问题进行剖析，针对性地提出考古类型学知识表示策略；其次，结合考古学的语料特征和类型学逻辑，对出土陶器的器形描述按照属种关系、整部关系两种维度分解；随后，对陶器的部分和类型逐层提出知识表示方案与特征向量表达式，进而构建出土陶器器形的知识表示模型；接下来，在知识表示模型的基础上，揭示基于条件等价映射的考古类型学本体扩展方法，实现考古类型学语义关联构建；最后，以青海柳湾的两件陶器为例，展示应用本文方法进行类型学知识表示的形式和过程，及其本体图形可视化效果。[结果/结论] 以陶器器形为例对出土文物从考古类型学视角下的知识表示和语义关联构建是数字人文研究在考古学领域的一种新的尝试，可以为类型学研究由依靠经验向依靠数据的转变提供技术支持。

关键词：考古类型学 出土陶器 知识表示 语义关联构建
分类号：G255
DOI：10.13266/j.issn.0252-3116.2022.12.009

1 引言

学科性是数字人文三个维度中的核心特质，如何从传统人文研究中思考数字人文的发展路径，以及数字技术与人文研究进一步融合发展的可能是数字人文工作者必须深入思考的问题[1]。考古学是数字人文研究的重要领域，目前考古学领域的数字人文研究在资料整理与登记[2]、知识服务平台建设[3]、GIS数字地图和场景复原领域[4]已取得了一定的研究成果，但是真正能深入考古学学科研究，采用数据科学帮助考古工作者进行专业分析的部分较为薄弱，这是当前考古学与数字人文领域的结合还未及深入所致。

地层学（stratigraphy）和类型学（typology）是考古学的两个方法论基础，前者用于获取考古资料，后者用于归纳和分析考古资料。两者结合是考古学时间序列构建和文化演进规律分析的必要条件[5]。传统的出土文物描述方式多用于在单一考古学文化内进行垂直的类型学研究，例如对某个遗址出土的某种典型遗存进行类型学分析，结合地层学搭建的时间序列归纳该遗址所代表的考古学文化随时间的演化规律，以此对相关的考古学文化进行分期、分段。上述工作主要是依靠考古学者的先验知识完成的，其所依据的定量方法主要来自统计学，但是由于缺乏成熟的样本检索系统和相似度分析方法，对考古资料的整理在应用中也未能深入到语义和知识层面，在相关考古学研究中难免会因知识表达方式受限而面临选取样本的覆盖率问题、选定样本的代表性争议以及样本分析的效率低下等问题。

虽然对出土文物的描述有约定俗成的术语系统雏形，但在至今仍未推行控制词表的考古学领域，无可避免地因语言表述的随机性，产生了一系列后续的语义问题，有时甚至在同一部文献中也会出现相应的语义混乱现象。例如，对于陶器中的“豆”，一般认为它从上到下由“盘”“柄”“足”三部分组成，但对于豆的“柄”，在有的文献中习惯称之为“把”，使之与有握把的其他陶器类型的“把”出现描述混淆，但显然不同语境中的这两个“把”所指称的在形态和功能上都是不同的事物。这种混乱状态造成了对跨文献单元出土文

* 本文系国家社会科学基金重大项目“新时代我国文献信息资源保障体系重构研究”（项目编号：19ZDA345）研究成果之一。
作者简介：韩牧哲，博士研究生；高劲松，教授，博士生导师，通信作者，E-mail：jsgao@mail.ccnu.edu.cn；李钰，讲师，博士研究生。
收稿日期：2021-09-21 修回日期：2022-03-13 本文起止页码：92-107 本文责任编辑：徐健

物的描述普遍存在差异性，且使得一些考古工作者在原始资料整理过程中无法参考和复用已有的描述规范，进而影响了考古学知识再生产的效率并可能使混乱状态进一步加剧。为此，俞伟超指出在类型学实践中"有许多器形，不知叫什么名称才准确"的难题，他认为在考古学跨文化研究中进行器物分析时，同名异器无疑会引起概念上的混乱，但是，想要将所有考古学文化中的器名统一，在很长时间内都不可能实现[5]，而从目前的考古学实践中来看，上述情况仍无显著改观。本文尝试面向考古类型学研究，提出一种适用于出土文物特征描述的知识表示模型，以对实物特征分解建模的形式打破考古学分类问题带来的语义壁垒，以实现对考古学知识的深层次、细粒度表示的功能。

2 相关研究

目前面向考古类型学知识表示和语义关联构建的研究仅见一例，A. Bruno 等以伊斯兰考古中的照明器具的分类问题为例，构建了相应的任务本体模型，这是对考古类型学进行语义组织的一次新的尝试[6]。但是文章研究的主题及其提出的词表和关联关系均无法复用到国内的考古学知识表示实践中，且欧洲伊斯兰考古的类型学基础与中国考古学存在很大差异。关于文物的元数据描述规范和本体框架大多是由以博物馆和博物馆联盟为代表的典藏机构发起并提出的，常用的专门元数据规范包括 CIDOC 概念参考模型（CIDOC conception reference model，CIDOC CRM）[7]，艺术品描述目录 CDWA（Categories for the description of works of art，CDWA）[8]，视觉资源核心目录（Visual resources association core categories，VRA Core Categories）[9]等，旨在为文化资源的典藏、陈设、共享提供标准规范和一致的信息描述框架。中国国家文物局和北京大学起草的文物保护行业的元数据编制规范体系中，涵盖了各类文物的元数据标准及著录规则，该标准的征求意见稿已经在网上公开发布[10]。上述知识表示方案可以为文物保护行业和典藏数据库建设提供可靠的资源采集和整理框架，并能在相应领域提供术语支持。不过，由于考古学领域对文物描述的目的、方式、细致程度和着眼点都与文物保护行业有一定差异，从文化遗产保护角度制定的元数据规范和本体框架普遍较为笼统，且针对性不强，其所揭示的知识粒度很难在应用中满足考古类型学分析和跨文化研究时的具体需求。目前，与考古学相结合的数字人文研究很少能深入其研究进程中真正的难点，对考古类型学的知识表示就属于数字人文真正深入考古学研究层面需要面对的关键难题之一。

本文计划提出的知识表示方案旨在为考古类型学研究提供技术支持，这要求在对文物本身进行描述时不仅要考虑其作为一件事物的整体表达，还要按照类型学的思维对其进行解构。对出土文物的类型学解构和对各个部分的知识表示，与工业设计领域的计算机辅助的装配制造问题有一定的相似性，知识表示方法和知识库建设已经在农机[11]、钢铁[12]、飞机装配制造[13－14]领域取得了显著的研究和应用成果，其中关于本体建模[15]以及在此基础上的信息检索和知识推送[16]方面的相关研究，对本文的方法体系构建具有一定的借鉴作用。在进一步开展研究之前，本文需要先将考古类型学的基本方法和思想进行简要阐释，剖析传统方法在数字人文进程中带来的主要问题，并明确类型学知识表示所欲达到的具体目标。

3 考古分类问题知识表示现状分析

考古类型学是考古学的基本方法论之一，类型学中最基本的问题是分类问题，由于从田野中获取的原始资料大量且细碎，考古学工作者不可能对每一件器物进行逐一描述，于是就需要进行分类，通过以不同层次分出的不同类别进行合理组织，最终可以构成分类的系统（hierarchy）[17]。分类思想伴随着中国现代考古学的诞生已经有百年历史，虽然方法几经变化，但思想是一脉相承的。它经常与"排队"（序列排布）思想组合使用，用于揭示时序性的发展规律问题。分类方法最常用于对陶器、青铜器和石器等文物的表述中，以及对墓葬、居址等遗迹形态的表述中。其中对陶器类文物的分类问题几乎是史前时代考古资料分析中最为普遍而重要的问题。

3.1 考古分类方法的知识推理过程分析

从科学研究方法来看，考古类型学主要使用的方法是归纳推理（inductive reasoning），其一般定义是从特殊到普遍的推理，是一种或然性推理[18]。在考古类型学中，单件文物本身或其可被直接观察记录到的其他特征个案，都属于归纳推理定义中的"特殊"；而归纳的作用则是从这些"特殊"的文物实体中寻找共性和差异性特征，在将具有较多相同特征的文物实体归结为类，形成指代一类文物的"普通"性表述；这种具有普通性表述的符号化呈现即为类型。由于考古发掘行为本身可以视为一种抽样式的信息采集，因此，对于某一遗址陶器器形的统计和归纳属于不完全归纳推理

活动,推理活动所产生的知识类型也属于不完备经验知识。

以新疆和静察吾呼四号墓地出土的陶釜为例展示分类法的知识推理规则,如图 1 所示:

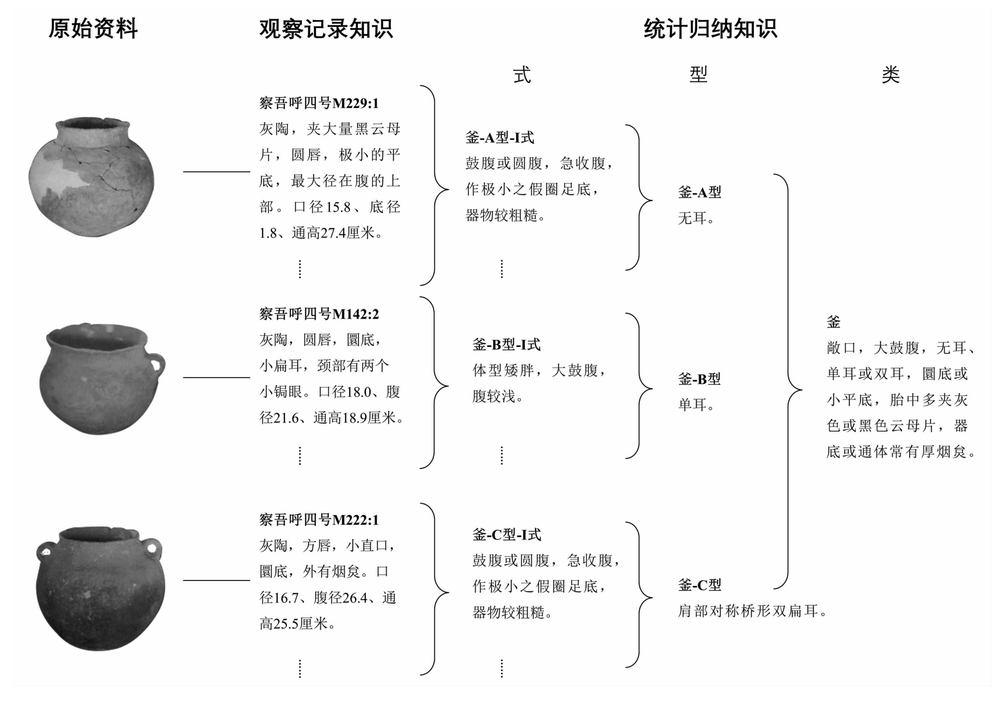


图 1 考古分类法的知识推理示例

由图 1 可见,察吾呼四号墓地出土的三件陶器在未经整理之前,属于原始资料。考古工作者通过观察对其进行描述之后,记录下来的知识,与单件文物一一对应。但是,仅在四号墓地,陶釜就出土了 64 件,由于考古报告的篇幅有限,不可能对存在批量性的陶器进行逐一描述,于是,就需要以归纳推理方法对具有相似性的文物进行归类,逐层总结出文物的类、型、式,从分类结果来看,文物分类描述中细节知识单元是按照式、型、类的顺序逐层减少的,其思维是逐层过滤掉文物中所包含的特异性,而保留文物中的共同性,最终的目的是以数量可控的类型学标记对该遗址或更大范围内出现的相应文物进行表示。不过,考古工作者对类、型、式的归纳可能并非顺序进行,而可能根据经验灵活运用向上归纳和向下细分两个方向,其基本思想仍是不完全归纳推理思想,为与此前直接观察记录所得的知识进行区分,归纳推理活动获得的知识本文称为统计归纳知识。

3.2 考古分类方法现存的问题分析

上述经典的分类思想,在以传统方法按照考古工作者的经验知识进行展开时,常存在或会带来以下问题:

3.2.1 分类所依据的标准较为随意

分类标准对应的是知识表示的维度,它是实现统一知识表示的结构基础,但是在考古类型学实践中,真正用来分类的标准在选取上非常灵活甚至可谓随意。例如,图 1 例中在同一层次的知识归纳中,所选用的特征知识单元就不统一,有的言及整体形状、有的提到局部特征和形状、有的则涉及尺寸或重量的比较问题;另外,除器形之外,纹饰类型、位置、图案和材质特征等都经常用于考古类型学描述,且没有固定的先后顺序和表述层次,着墨也繁简不一,这使得在进行考古类型学知识组织时,很难找到统一的知识结构框架。

3.2.2 描述所使用的术语体系混乱

虽然可以复用 CIDOC CRM 本体族中的概念以及其他相关元数据模型对考古学领域知识进行表示和组织,但是目前没有一种现存的术语词表可以对考古类型学中涉及的概念进行妥善处理,在考古学实践过程中,术语体系混乱的问题也经常出现,这一现象在前文中已做示例;类型学描述中的另一个重要问题是模糊表述问题,考古类型学描述中大量使用表示模糊语义的相对概念,如大小、高低、粗细等,且可能用到不同层次的比较词汇,这是广泛存在于人文社会科学领域的不完备经验知识的特性,对这类描述进行知识组织和表示的难度较大。

3.2.3 类型学符号标记缺乏共享融合方案

目前中国考古学使用的类型学符号标记方法是有

一定规范的，通常是以“类别名（汉字）＋型别名（大写拉丁字母）＋式别名（大写罗马数字）”的形态呈现，有时，考古工作者会根据需求对层次进行删减或扩充，且并不一定追求符号标记系统的规范性。由于考古报告通常是以遗址为单元组织的，因此，考古工作者会在同一份报告中保证类型学符号标记的统一性，以方便进一步的统计和分析。例如，图1例中察吾呼四号墓地墓葬登记表[19]的M229中即可直接以“釜AI”为文物M229:1进行标记。但是，对不同遗址甚至对同一遗址不同的发掘或整理活动批次下所产生的成果中，这套符号标记体系都难以互通，例如，察吾呼遗址的四号墓地和一号墓地都是由新疆考古研究所在1985－1989年间相继发掘，且完善的发掘报告以专著形式一同出版，但是即便这两座墓地中的陶器类型学标记也有少量不一致处，例如，四号墓地中的“翻耳杯A型”“翻耳杯B型”分别对应一号墓地的“翻耳杯Ⅱ式”“翻耳杯Ⅰ式”，这造成了两座墓地的墓葬登记表中的随葬品等级信息也未能统一，为后续的研究者在资料使用和统计时带来障碍。考古分类方法的应用通常以一个遗址或一个文化为单位垂直进行，在跨遗址或跨文化研究中，原本的类型学符号由于缺乏融合方案，往往直接被弃之不用，成为“一次性的”符号体系。在新的研究过程中，考古学研究者往往要重新根据少量描述全面的原始资料，按照自己的知识结构进行重新归纳和类型划分，严重影响了工作效率。综合以上三点可以看出，传统的考古分类方法中的知识组织结构、术语描述系统、符号表示系统都存在较明显的缺陷，它们在一起使得原本为了方便知识交流和理解而创造的类型学符号标记系统，在数字环境下形成了一种阻碍知识共享和重用、不利于刻画和还原物理事物和事实现象的语义壁垒。例如，对于察吾呼四号墓地M95中登记的“釜CI”，由于在报告中未做单独展开，该件文物的信息就只能通过对该类型陶釜的描述演绎推理获得，除明确可知其具有双扁耳以外，对于底形、腹形、唇形等细节特征的掌握都是非确定的。

而且，目前基于传统分类方法的器物描述现状，非常不利于后续数字人文方法的引入，现有的类型学符号标记系统会掩盖作为特征项的细节知识，使得基于此的知识应用效果大打折扣。类型学符号标记几乎不包含任何语义信息，此处本文继续以一场景示例加以说明，其所使用的原始语料摘录见表1。

假设在察吾呼文化分期研究中，要对表1所示的两座墓葬遗迹的相似度进行计算，以两座墓葬的随葬

表1　察吾呼IV－M122和I－M275墓葬登记信息摘录

遗址	遗迹	随葬文物
察吾呼四号墓地	M122	曲腹杯I，翻耳杯B，铜锥I
察吾呼一号墓地	M275	翻耳杯I，釜AI，砾石I

文物信息来看，其类型学标记无一相同，且单凭文物类名也无法提供可供比对的语义信息，于是，即便有相应的语义距离测量方法参与，也无法得出可靠结果。

不过，笔者调查后将四号墓地和一号墓地的“翻耳杯”的分类信息摘录如图2，从图中的描述和示意图可见，四号墓地的“翻耳杯B”与一号墓地的“翻耳杯I”应属同类。在此认知的基础上，在对表1中的两座墓葬进行对比，即可对二者在出土文物上的相似度做出较为客观的评估。因此，如何突破当前考古分类方法造成的语义壁垒，是展开进一步知识组织和应用的前提。

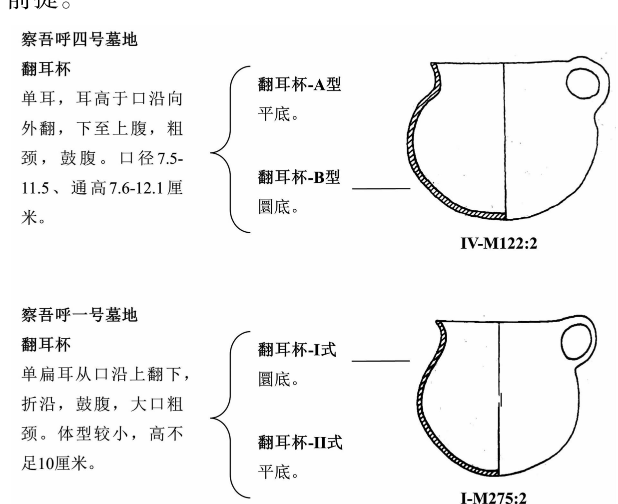


图2　察吾呼四号墓地和一号墓地的翻耳杯类型信息摘录

3.3　考古类型学知识表示的策略

由以上分析可知，当前考古分类方法上的主要问题在于其知识结构、术语体系、符号体系的不统一，很大程度上削减了其所承载语义信息的可理解性，使得其有效应用范围非常狭隘。要解决上述问题，需要针对考古类型学建立统一的知识表示方法，并预设相应的形式化方案，使之能与领域本体以映射方法集成，进而才能实现类型学知识与考古学领域知识的融合。

面向考古类型学的知识表示方案的第一步，是从原始语料中寻找可用于描述相应事物的通用语。进而以通用语表达的知识单元为基本单位，以事物内在的逻辑关系为依据建立知识表示结构模型。

其中，通用语主要分为两种：一种是人可通过常识理解的一般概念，不需要任何专业背景和认知，例如对

"碗""盘""口""底""大""小""圆""扁"等，这类概念及其在某种语言中的称谓已经约定俗成，极少会造成语义理解障碍；另一种是目标领域内的通用语，这些概念的适用范围通常是领域内部，可能需要基础的专业认知，例如"簋""甗""圈足""流""侈""敛"等，这些概念在考古学和文物与博物馆学领域也是通用的，不需要作其他解释和限定。通过示例可以发现，通用语描述的都是非常单一的特征，"碗""盘""簋""甗"等都是陶器的类别名称，"口""底""圈足""流"则是陶器的不同局部称谓，这些名词属性的通用语往往是器物或特征主体的代称，在语义网中对应的即实体(entities)；而"大""小""圆""扁""侈""敛"等则属于形容词性的修饰语，是对器物或特征主体的修饰，在语义网中对应的是属性(properties)。纵观整个类型学描述，可知绝大多数陶器类型都是按照以上述通用语组合而成的定中短语(定语 adj. + 中心词 n.)表示的，其结构具有一定的稳定性，说明考古类型学知识是具备知识表示基础的，一般而言，这种知识结构可以采用特征向量(eigenvector)进行转换和表示。

针对考古类型学的知识表示是否能够打破前节分析的语义壁垒，促进知识共享和复用，是本文研究的意义所在。从前文关于"翻耳杯"的案例可知，传统的类型学符号标记系统难以对相关器形的特征细节进行直接表示，这使得其所包含的有效语义信息极少；不过，假如以其描述中的通用语短语解构成独立的词和字，按照其原有的描述逻辑重新组织，建立知识表示模型之后，即可通过知识表示模型对某一文物所关联的知识单元及其关联关系进行结构化存储，为后续的知识应用和其他数字人文方法的引进奠定基础。接下来，本文将结合实际案例对面向考古类型学的知识表示方法进行分析。

4 数据采集和整理方案

由于出土文物的门类庞杂，本文选取在考古类型学中应用最广泛的陶器作为研究对象，样本范围被限定为器形完整的陶制容器及其器盖等必须组合部分，如陶器残片，纺轮、陶球一类的陶制工具，以及陶俑、支座等不属于或不确定为容器的样品将被排除在外。

本文研究主要面向的是考古学领域的研究者及其学科研究过程，因此在知识获取阶段仍然以在学科领域内具有权威性的文献作为组织语料库的数据源，结合一定的自然语言处理方法从中提取和指定术语，并基于真实样本定义属性和关系。考古类型学主要应用于史前史研究，新石器时代也是陶器应用于考古类型学研究最为重要的阶段，综合考虑数据源的权威性和代表性后，本文选取中国社会科学院考古所编著的《中国考古学·新石器时代卷》[20]作为文献数据源进行语料采集，部分线图和案例资料以青海[21]和新疆[19]的考古报告补充。

4.1 出土陶器原始术语词表的组织

新石器时代遗址中出土的陶器一般是文化特征的主要标识物，文献数据源中有关出土文物的描述主要在"文化特征"中阐述。本文的语料库采集和处理策略分为三个步骤：①人工遍历整部文献的文字内容，将正文中描述陶器的部分以句为单位摘录，并以序号形式存储；②对采集的自然语句进行拆分，仅保留与陶器描述有关的词汇和短语，这项工作需要对部分句子结构进行重组或语义补全，以保留相关词汇的关联信息；③对拆分后得到的词汇和短语进行进一步拆分，整理、清洗、去重后组成术语词表，以叙词表形式保留可以推导出的语义关系，作为后续语义关联构建的基础。具体过程示例见图 3。

考古学领域对一件陶器类文物的描述，通常涉及器类、器形、质地、制作工艺、陶衣、彩绘纹样等多个方面，语料采集和原始术语提取的过程中，将出现的术语按照其所描述的内容结合语境进行归类，并将其中的属种关系和整部关系以叙词表的形式组织。由于单一案例中的样本量有限，一些在整理前期不易被归类的术语，将在整理过程中随样本量增加重新归类。图 3 所示的原始语料 215.2.1 中的 6 个句子，已涵盖了 5 个方面共计 61 条术语，完整版的术语词表所涉及的方面和层级会更多且更复杂。因此，划分层次对于清晰的知识表示意义重大，对于陶器类文物实例的表达，理论上可以简单划分为器形、质地、工艺、装饰 4 个描述维度，倘若考虑到考古学文化层面的关联，还可以融合文化、年代、共出物等关联维度，从而使得对于陶器本身的表达可以融入到已有的语义知识网络中被重用。本文选取类型学中最常用于分析的"器形"作为主要对象进行展开。按照上述方法将数据源中的所有原始语料进行词表组织，从中抽取有关"器形"的语料部分，即为下一步进行特征提取的素材库。根据"器形"素材库所包含的内容，可以将其中经过拆分的词语划分为 5 种类型，分别为器形术语、部分术语、相对修饰、描述修饰、定量修饰，这 5 种词汇所对应的概念中，器形术语和部分术语是用于描述目标文物整体和局部特征的概念，其他的 3 种修饰语则是对上述概念的相关特征进行修饰的属性。

韩牧哲，高劲松，李钰.面向考古类型学的出土陶器器形的知识表示与语义关联构建[J].图书情报工作,2022,66(12):92-107.

原始语料（215.2.1）

S1：陶器为手制，质地多为夹砂或泥质红陶，有少量黑、灰色陶器。
S2：器形主要有圜底和小平底钵与盆、深腹盆、细颈大腹壶、小口尖底瓶、深腹罐，多圜底、平底和尖底器，少圈足器，无三足器。
S3：晚期出现葫芦瓶、带盖平底小罐和高领罐（图4-2）。
S4：纹饰有绳纹、细绳纹、弦纹、锥刺纹和黑彩图案。
S5：彩陶纹样有宽带纹、三角、折线网纹等几何纹和鱼纹、人面纹、鹿纹、鸟纹等象生图案，常见内彩。
S6：有些钵盆类陶器见有不同的刻符。

注：取自《中国考古学·新石器时代卷》第215页第二自然段。

自然语句拆分和术语抽取

S1：手制；质地；夹砂；泥质；红陶；黑陶；灰陶
S2：器形；圜底；底；小平底；钵；盆；圜底钵；圜底盆；小平底钵；小平底盆；深腹；腹；深腹盆；细颈；颈；大腹；壶；细颈大腹壶；小口；口；尖底；瓶；小口尖底瓶；罐；深腹罐；平底；圈足；足；三足
S3：葫芦瓶；盖；小罐；带盖平底小罐；高领；领；高领罐
S4：纹饰；绳纹；细绳纹；弦纹；锥刺纹；黑彩
S5：彩陶纹样；宽带纹；三角纹；折线网纹；几何纹；鱼纹；人面纹；鹿纹；鸟纹；象生图案；内彩
S6：刻符

注：“A”为词义补全部分。

原始术语词表组织

制作工艺
　　NTG　手制
质地
　　NTG1　胎质
　　NTG2　　夹砂
　　　　　　泥质
　　NTG1　胎色
　　NTG2　　红陶
　　　　　　黑陶
　　　　　　灰陶
器形
　　NTG1　钵
　　NTG2　　圜底钵
　　　　　　小平底钵
　　NTG1　盆
　　NTG2　　圜底盆
　　小平底盆
　　深腹盆
　　NTG1　壶
　　NTG2　　大腹细颈壶
　　NTG1　瓶
　　NTG2　　小口尖底瓶
　　　　　　葫芦瓶

　　NTG1　罐
　　NTG2　　小罐
　　NTG3　　　带盖平底小罐
　　NTG2　　深腹罐
　　　　　　高领罐
　　NTP1　底
　　NTP2　　圜底
　　　　　　尖底
　　　　　　平底
　　NTP3　　　小平底
　　NTP1　腹
　　NTP2　　深腹
　　　　　　大腹
　　NTP1　颈
　　NTP2　　细颈
　　NTP1　口
　　NTP2　　小口
　　NTP1　足
　　NTP2　　圈足
　　　　　　三足
　　NTP1　领
　　NTP2　　高领
　　NTP1　盖

纹饰
　　NTG1　绳纹
　　NTG2　　细绳纹
　　NTG1　弦纹
　　　　　锥刺纹
彩陶纹样
　　NTG1　几何纹
　　NTG2　　宽带纹
　　　　　　三角纹
　　　　　　折线网纹
　　NTG1　象生图案
　　NTG2　　鱼纹
　　　　　　人面纹
　　　　　　鹿纹
　　　　　　鸟纹
未归类
　　NT　黑彩
　　　　内彩
　　　　刻符

注：“A”为词义补全部分；
“NT”为下位词，
“NTG”为属种关系下位词，
“NTP”为整部关系下位词，
阿拉伯数字表示层级。

图3　原始术语词表的组织过程示例

4.2　出土陶器器形描述的维度

传统的应用于博物馆学中的元数据，通常对于文物的描述不会深入到部分层面，即便涉及对组成部分的描述，也会被混杂在器形描述中。然而，从考古类型学角度对器物进行分析，对部分特征的描述有时会成为区分陶器时序演化的非常关键的转折点；而且，除了对陶器形态的描述性修饰以外，一些具有比较属性的相对特征和定量特征在分析过程中同样不容忽视。因此，本文将属种关系和整部关系作为有关陶器器形的两个主要维度，属种维度将单件陶器视为整体，并以其整体形制所代表的实用功能作为陶器类别的划分依据，但在类别确定的基础上所做进一步的型别和式别划分，则主要考虑陶器的部分特征；整部维度将陶器整体进行结构拆分，划分出若干组成陶器的功能部分，并进一步对每个部分可能存在的定性和定量修饰属性进行排列组合，进而实现对该部分精准描述的目的。

4.2.1　陶器器形的属种关系

目前的考古类型学是一种高度依赖经验、特征点不统一、知识粒度需求差异较大的复杂领域，类型学的主要工作是定义陶器器形的属种关系，其划分逻辑自上而下分别为：类别、型别、式别，由抽象到具体，并最终能与代表具体文物的实例关联。

一般来说，陶器的类别（variety）由先验知识预判完成，其所依据的主要是某类陶器实用功能，如炊器、饮器、食器、盛贮器、汲水器等。特定的器形，通常存在约定俗成的类别名，如一般人通过常识就可辨识生活中的“碗”“盘”“瓶”“杯”等不同类的器皿，对出土陶器类别的分辨主要也依靠经验，少数退出实用领域的古老器类会参照文献典籍中的称谓和定义，如“鼎”“鬲”“豆”“鬶”等。陶器类别的辨识特征可定义为“类名主体”，对类名主体的提取主要依靠人工标注进行，人工提取的结果即表现为顶层的器形术语。

确定类别之后，对同类陶器的型别划分，主要依靠一些特征性的部分元素，如对于考古学中的“壶”进行分型，可根据其局部特征将其划分为“腹耳壶”“贯耳壶”“带流壶”等，其中具有特征性的部分术语如“腹耳”“贯耳”“流”等即可称为“型别特征”。

如果在一个遗址或一个文化区域发现了属于同一

类型但式样存在差异的陶器样品，则需要进一步划分式别，式别的划分没有统一标准，而且有时会考虑彩绘纹样、装饰风格、陶胎质色等器形以外的相关因素。只考虑器形的话，主要参照的是添加了各种修饰限定的器形术语或部分术语，这些特征可以统称为“式别特征”。对出土陶器的类型学描述示例如图 4 所示：

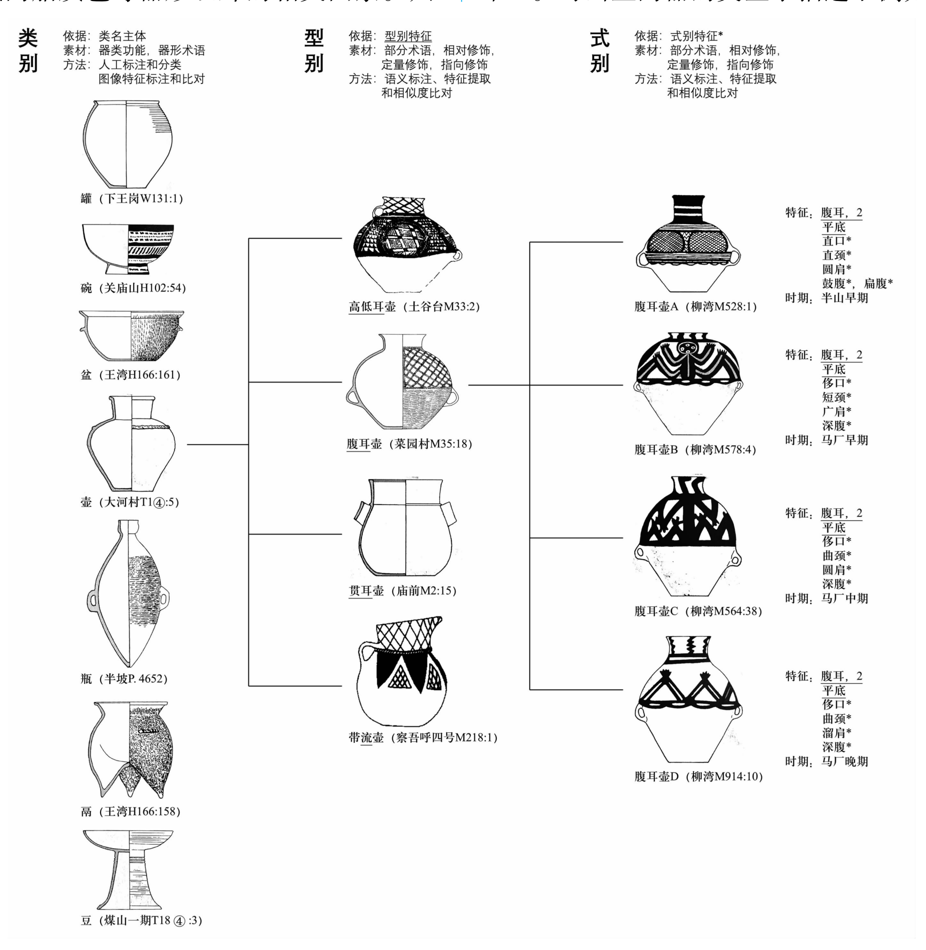


图 4 出土陶器的考古类型学划分逻辑示例

注：腹耳壶分式仅为示例，进行了适当简化，并非原考古报告中的类型划分

4.2.2 陶器器形的整部关系

考古类型学很少专门讨论陶器的整部关系，但是在类型学分析中，陶器的部分特征是主要的参考要素。例如，当对同一遗址出土的单一器类“腹耳壶”进行式别划分时，宜将其统一的型别特征“腹耳”“平底”作为固定参数，其特征点主要是口、颈、肩、腹的形态，用于比对知识主要来自“部分术语”的“相对修饰”层面；而如果对某一遗址不同年代分期或某一考古学文化的不同类型的特征进行归纳，则可能从器类、乃至跨器类的部分特征进行展开，如新疆的察吾呼文化以带流器为典型，这里对“部分术语”中“流”的考察则会囊括带流杯、带流罐、带流壶，是跨器类的。

目前，在考古学领域尚未见到对陶器部分的分类研究和术语规范工作，本文也只能依靠选定文献数据源中的信息进行归纳。为盛放、贮存其内容物，陶制容器通常具有口上底下的结构，所以本文将陶器的各个部分按照其结构位置划分为上侧部分（P_H）、中间部分（P_M）、下侧部分（P_B）和附件部分（P_A），各部分的划分方式和示例见图 5。

从图 5 的上中下各区域的结构划分方法延伸，还可以按照部分的上下结构进行扩展标记，编号规则以对应部分位置越高，序号越小，如中间部分的肩通常在腹的上方，则样品“壶（大河村 T1④:5）”的“广肩”和“腹・UD”可分别标记为 P_{M1} 和 P_{M2}。

各个部分术语还能分离出很多经过修饰限定的子类，从文献数据源中分离出的术语统计可知，修饰词主要分为三种类型，分别为描述修饰、相对修饰和定量修饰。描述修饰是对部分特征的定性描述，如“平底”“圜底”“腹耳”“袋足”等；相对修饰是具有比较级或模糊定量作用的修饰词，多为修饰形态的词对，如“大/

小、凹/凸、宽/窄、侈/敛”等；定量修饰是相关部分的数量属性，如双耳器所包含“耳”的数量为2，三足器所包含“足”的数量为3，除表示数量以外，尺寸、比例、角度、相对位置坐标和其他有意义的统计值均有可能是定量修饰的备选项，数据类型也可能更为灵活多样，关于属性的细分属于考古学者的研究范畴，本文不再赘述。

上侧部分(P_H)：口，唇，沿，流(口)[1]，颈……
中间部分(P_M)：肩，腹，壁，流(管)[2]，盘(部分)[3]……
下侧部分(P_B)：底，假圈足(底)，(空)足，(实)足，圈足，柄[4]……
附件部分(P_A)：耳，鼻，鋬，钮，把，提梁，脊，盖……

注：1.流(口)，特指与上侧口沿一体的流，常见于带流杯或匜。
2.流(管)，特指与中间肩部或腹部连接的管状流，常见于盉。
3.盘(部分)，特指完整陶器中作为盛物盘的部分，常见于豆。
4.柄，指连接、支撑盛物盘的长柄形支座或底足，常见于高柄杯、豆，经常被混淆称为“把”，需与附件部分中连接与器物侧面的“把”区分。

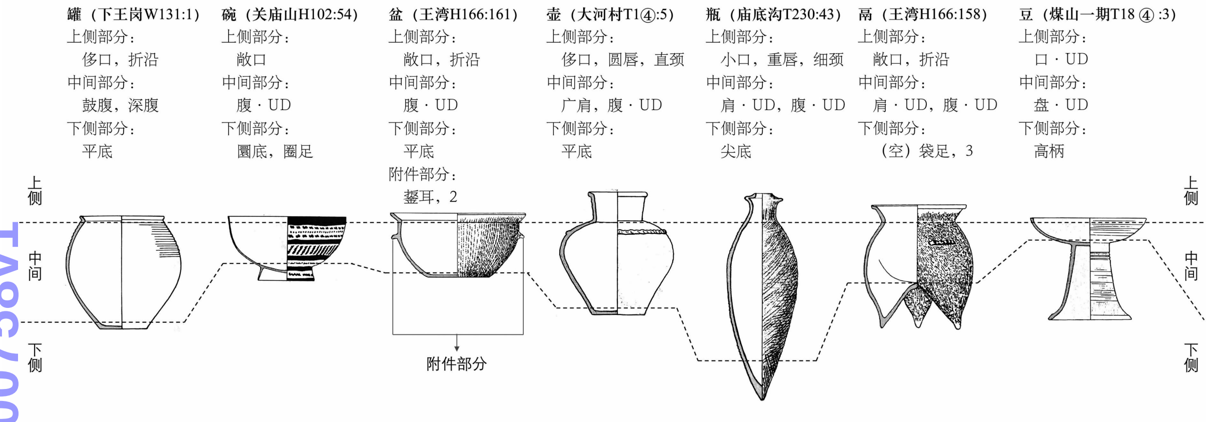


图5　按照结构位置对陶器部分的分类与示例

5　出土陶器器形的知识表示模型

在明确出土陶器器形的属种关系和整部关系两种描述维度之后，需要进一步对陶器器形进行知识表示，此处仍须遵循考古类型学的逻辑进行逐层表示，逐步细化后最终提出整体的陶器器形知识表示模型。

5.1　对陶器部分的知识表示

由上文可知，陶器的部分是陶器类型划分的重要参考因素，因此，在形式化表示过程中，可以将部分概念作为陶器类型的对象属性（object properties）。由于部分还可以按照三种修饰术语的组合进行更细致的划分，所以需要将三种作为定语的修饰术语表示为每个部分的数据属性（data properties）。划分后，陶器部分的知识表示需要包含该部分的结构位置，以及针对该陶器部分的三种修饰属性，修饰属性的不同组合方式均可作为陶器部分的细分类标准。对陶器部分的描述仅可设置三个层次，顶层是部分的结构位置，即标记该部分位于陶器的上中下侧或为附加部分；中层是未指定任何修饰属性的默认部分（P_m），例如笼统地“口”“腹”“底”“耳”等，图4中带有“UD”标记的部分均未指定修饰属性；由此向下添加修饰属性做出任何细分的部分均为具体部分（P_{mn}），底层部分的描述需包含可说明结构位置（L）及其上层默认部分的关联信息和描述修饰（F_D）、相对修饰（F_R）和定量修饰（F_C）三种属性，其公式表达为：

$$P_{mn}=\{R_L,R_{P_m},F_D,F_R,F_C\}.\quad L\in(H,M,B,A)$$

公式(1)

式中，P_{mn}代表具体部分；R_L代表该具体部分的结构位置，其取值可以区分该部分属于上侧（H），中间（M），下侧（B）或附件（A）；R_{P_m}是该具体部分所属的默认部分；F_D，F_R，F_C是具体部分的三种属性，属性数量不做限定。

在实际的数据源中，三种修饰属性都存在未指定的情况，对未指定的属性应给予默认赋值，具体的赋值方式需要结合具体情况进行具体分析。

5.2　对陶器类型的知识表示

考古类型学的固有逻辑是本节接下来进行陶器类型划分和知识表达的主要依据，单纯考虑属种层次和陶器部分表示中的器形部分，可将出土陶器的类型划分逻辑简化表示，见图6。

从类型学角度对于一件出土陶器器形的描述要在划定类别的基础上，对陶器组成部分的特征进行逐层比对，随后按照组内相似、组间相异的原则进行型别、式别划分。不过，图6模型仅能展示陶器类型划分的

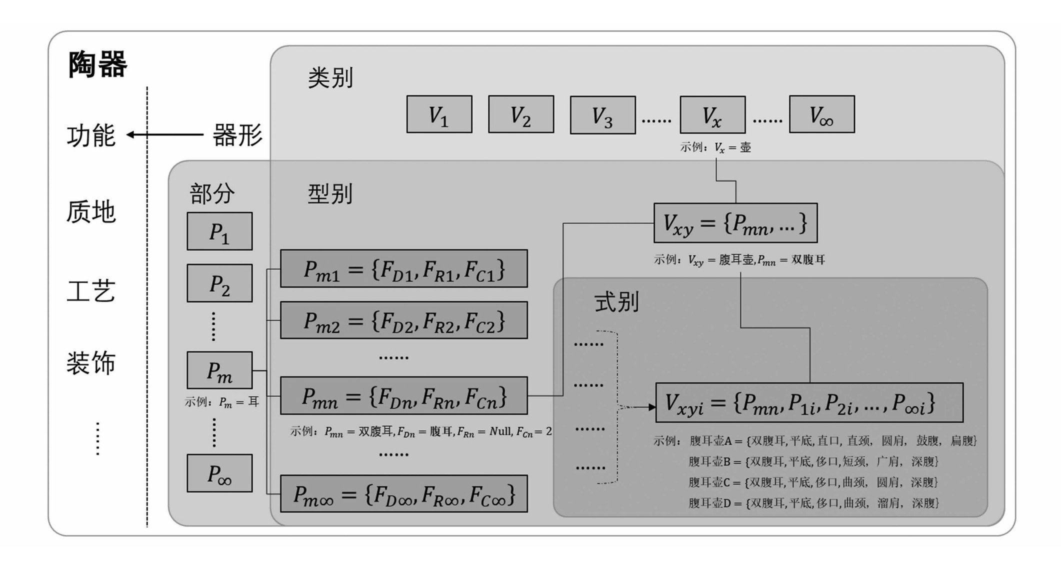


图 6　出土陶器的考古类型学划分逻辑模型

注：F_D 为描述修饰属性，类型 – string，值域 – 素材库 – 描述修饰词；F_R 为相对修饰属性，类型 – string，值域 – 素材库 – 相对修饰词；F_C 为定量修饰属性，类型 – integer，值域 – 自然数 – 整数

基本逻辑，但未能涵盖关联信息，也没有写入可供推理和计算的相关属性。接下来，将对陶器的器形描述进行逐层展开和规则细分，进而提出可供推理和计算的具体知识表示模型。

一些常见而固定的类别通常有特定的整体器形，器形进而可以分解出该类别必要包含或不包含某种部分，例如，陶器类“鼎”必有“底”，底部必有分裆的“实足”；而“实足”与“柄”“圈足”“空足”等部分概念集合存在互斥关系，这就使得“鼎”与上述部分之间也形成另外一种必然关系，即“鼎”不可能包含上述部分，反推可知包含上述部分的陶器类别不可能是“鼎”，这些公理可以在知识图谱应用中以语义推理的方式实现对一些陶器类别的划分。但是，能够通过语义推理确定陶器类型的可能性很小，对于很多可以包含同类部分，乃至功能和整体形状相似的样品而言，在数据规模可控的情况下，人工标注仍是当前最有效的类别划分方法。近百年前，梁思永[22]、李济[23]等学者就已经通过定性与定量结合的方法对陶器形状的人工标注进行过尝试，类似的思想如今仍可应用于未来可能实现的半自动或自动特征提取和语义标注过程中。此外，规格尺寸也是类型划分的参照依据之一，有些样品的形状和组成部分均较相似，但规格悬殊，这很可能直接反映其在古代的功能差异，在类别上自然也不宜归并。综上，对于出土陶器的器形表达，可以包含 4 个方面的信息，分别为关联信息（relations），组成部分（parts），外形轮廓（outline），规格尺寸（size）：

$$V = \{R, P, O, S\} \qquad \text{公式(2)}$$

5.2.1　对陶器类别的知识表示

对于类别的描述，以上 4 个方面所涵盖的信息较为笼统，其中关联信息的写入采取向上关联法，而类别是器形属种关系中的顶层术语，没有相关的上位类，在类别的相关信息中写入的是该类陶器的实用功能（R_{Usage}）。同一类别陶器的组成部分和结构会存在个体差异，因此，在类别表达式中很难写入具体的类，这里需要写入的是两种可以说明公理的限定性部分集合，分别为类别 V_x 必须包含的部分 P_{V_x} 和不可能包含的部分 $P_{\overline{V_x}}$。外形轮廓的描述方式有很多，基于二维或三维坐标系的图像特征标注方法在图像检索领域已有较为成熟的应用，但建模所需的成本较高，暂时没有引入的必要性，而且同类别陶器的形态和尺寸可能颇为悬殊，因此在类别表达式中不做赋值。陶器类别的表达式即可写作：

$$V_x = \{R_{Usage}, P_{V_x}, P_{\overline{V_x}}\}. \quad P_{V_x} \cap P_{\overline{V_x}} = \emptyset \qquad \text{公式(3)}$$

式中，V_x 代表某种陶器类别；R_{Usage} 表示该类陶器可能具有的实用功能；P_{V_x} 代表该类陶器在完整状态下必须包含的部分；$P_{\overline{V_x}}$ 代表该类陶器不可能包含的部分，P_{V_x} 和 $P_{\overline{V_x}}$ 的取值通常都是默认部分。

这种表达式本身还可以存储一定的语义信息，仍以“鼎”作示例，其类别表示模型见图 7。

5.2.2　对陶器型别和式别的知识表示

接下来对型别（V_{xy}）的描述中，关联信息可链接到其所属的类别信息的 URI，并继承其类别信息中已定义的属性。组成部分方面则重点写入该器形标志性的型别特征部分，型别特征部分取值范围为其所属类别

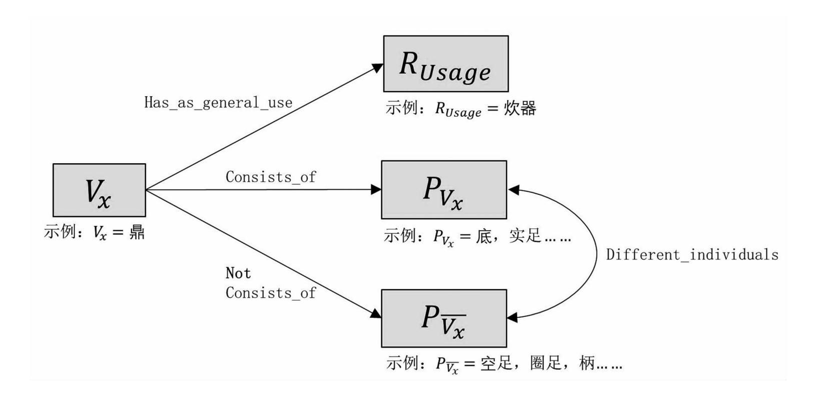


图7　陶器类别的知识表示模型

“不可能包含的部分”集合的绝对补集，记为$\overline{P_{\overline{V_x}}}$。同样，陶器的型别层次依然不足以实现对形态和尺寸的统计，因此在型别表达式中也不做赋值。陶器型别的表达式可写作：

$$V_{xy} = \{R_{V_x}, P_{V_{mn}}\}.\ P_{V_{mn}} \subset \overline{P_{\overline{V_x}}} \qquad 公式(4)$$

式中，V_{xy}代表某种陶器型别；R_{V_x}表示该型陶器的直接上位类，即其直属的类别；$P_{V_{mn}}$是其型别特征部分，该部分的取值范围为$\overline{P_{\overline{V_x}}}$。

式别(V_{xyi})表达式涵盖其所属的型别信息的*URI*并继承属性。组成部分方面需写入该式别陶器拥有的所有具体的特征部分。同一式别的陶器通常具有相似的形态，因此，在陶器的式别表达时需要描述其外形轮廓，图像检索领域的相关算法能够实现对样品整体二维轮廓和关键部分特征的提取和标注，进而通过对标注图像特征的相似度计算完成对陶器类别的匹配和推荐[24－25]，但是在数据量可控的情况下，人工标注和基于统计的描述在传统人文学科中的实用性更强。以统计方法描述陶器基本外形轮廓的手段有多种，因考古文献资料所展示的陶器图片大多为平视线图，且很多未标注原物尺寸，对其进行外形轮廓刻画需要通过比例描述，且只考虑其器形的主体部分，忽略附加部分。有两种方式可行性较高：一是切线定点法，基本思路是将陶器的底平齐于 *X* 轴，使底部连线的中点位于直角坐标系原点，对陶器外形轮廓做内外切线，记录每两条切线的交点坐标，通过坐标位置还原比对即可对陶器基本外形轮廓进行刻画；另一种是横切比例法，适用于主体形状对称的陶器，分别按照从上中下三侧选取其横截面直径最大和最小的两个部分，记录各部分直径，计算并写入不同部分直径的比值，可以大致判断出该陶器形状的基本走向，两种方法得到的外形轮廓描述分别为坐标集和数组，在相似度计算时需要选区的方法不同。同式陶器往往器形规格差距不大，可以写入代表性样品的规格尺寸或多件样品的尺寸均值作为参考。直接通过描述具体文物而得到的陶器实例的表示方法和式别基本相同，可将陶器式别和陶器实例的表达式写作：

$$V_{xyi} = \{R_{V_{xy}}, P_{V_{xyi}}, O_{V_{xyi}}, S_{V_{xyi}}\}.\ P_{V_{xyi}} \subseteq P_{mn} \qquad 公式(5)$$

式中，V_{xyi}代表陶器式别或实例；$R_{V_{xy}}$代表该式或该件陶器直接关联的上位类，即其直属的陶器型别(如未分类或不确定则为类别)；$P_{V_{xyi}}$是其式别特征部分，取值范围为具体部分；$O_{V_{xyi}}$为通过轮廓描述方法计算得到的器形轮廓信息；$S_{V_{xyi}}$为该式或该陶器的规格尺寸信息。

5.3　出土陶器器形的知识表示模型

综合已有的素材，可以进一步结合实例构建出土陶器的器形的知识表示模型，如图8所示。该模型可将从考古学资料中描述逐层抽取的陶器器形信息进行知识表示，除了整合现有数据库中的结构化数据以外，其给定的描述方法还能方便用户通过文献描述或直接对实物进行知识抽取。模型的结构层中，有关陶器类型和部分特征的术语均从考古学语料库中整理、归纳而得，最大限度地复用学科术语并还原考古学的表达方式，未来，实例层中收录的经过语义消歧和去重的术语能够以规范词表的形式向用户提供参考，以照片、线图等形式对关联术语进行直观展示，从而极大地方便考古学工作者的资料整理过程，并实现最大限度的术语重用和一致表达。

结构层为部分的三种修饰属性分别进行了取值示例，这三种属性的不同取值组合之后即可实现对任意陶器部分的精确刻画，且根据实际需要，通过赋权和权重控制还能够实现可满足不同情境的定制式知识表示。模型示例中对三种属性的设置都进行了简化，其中 F_D 和 F_R 两种定性修饰属性在确定之前通常需要进行语义对齐，F_R 的取值有时需要给定参考范围，并且定性属性在应用中可能存在多个取值同时存在的情况；

图书情报工作
第 66 卷 第 12 期 2022 年 6 月

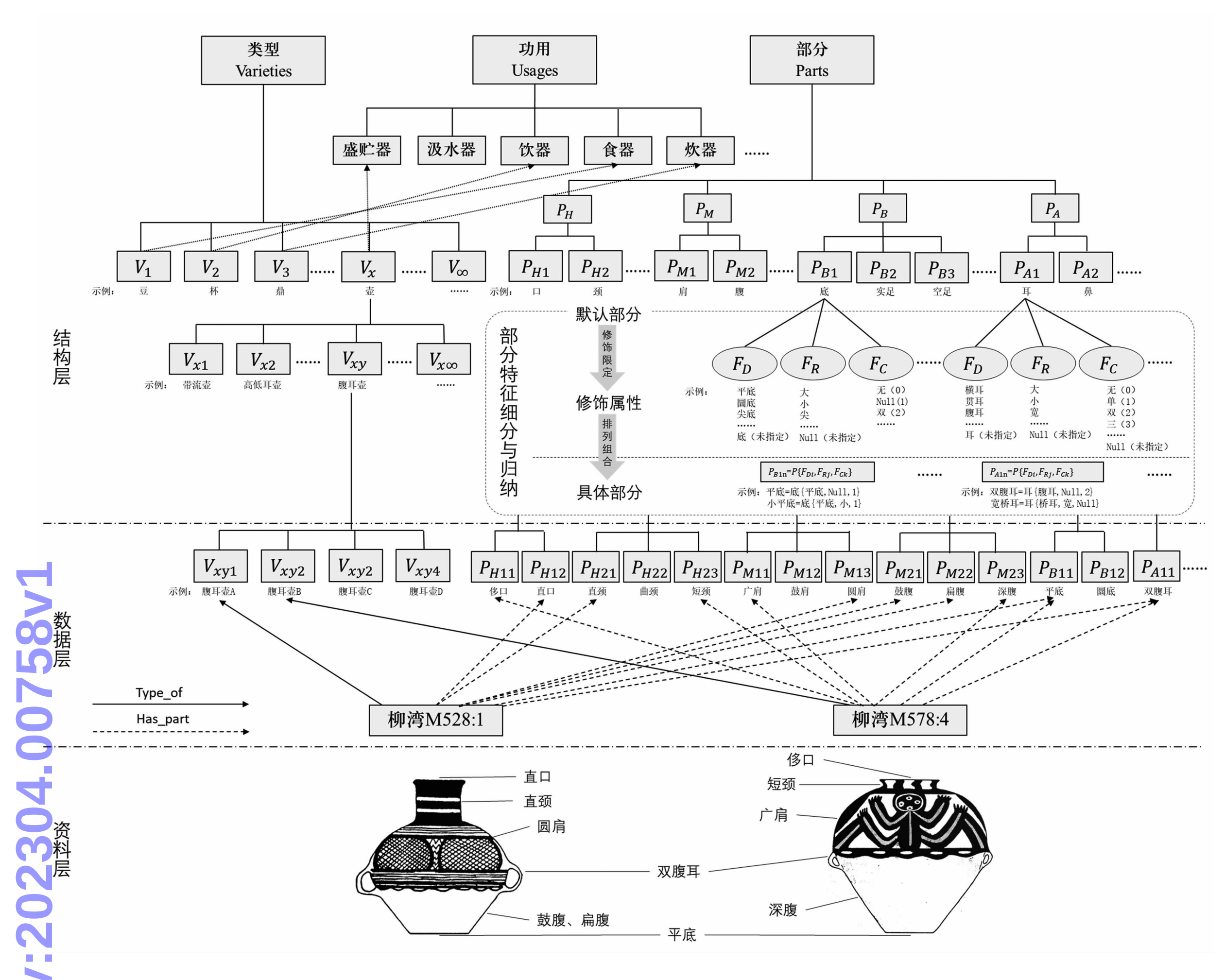


图 8　出土陶器的器形知识表示模型

示例中 F_C 仅以“数量”参数作示例，实际的取值方式可能非常丰富。

按照这种方式对出土陶器器形的知识表示，是进一步实现对形式化表示的基础。考古学分类问题中的对其他诸如陶器装饰纹样、胎质以及对青铜器、玉石器等文物的知识表示逻辑和实现方法与对陶器器形的表示大同小异，怎样将上述知识表示的结构逻辑本体化，使之能够实现与考古学领域知识图谱的语义集成是接下来研究的重点。

6　出土陶器器形知识的语义模型构建

前文提出的出土陶器器形的知识表示模型完成了对目标领域术语概念化和术语关系的表示工作，主要解决了“事物可描述”的过程。想要真正在应用层面实现知识组织和表示，需要进一步对其进行基于本体的形式化表示，从而完成“机器可实现”的功能，使本文提供的面向考古类型学的知识表示方法能够与领域知识图谱进行知识融合，实现更深层次的知识服务目标。

6.1　基于语义映射的考古类型学本体扩展方法

与一般的本体构建方案不同，本文并不打算构建独立的领域本体模型，而是尝试针对出土陶器的器形表示问题，提出一个可以与现有相关的领域本体实现集成的扩展方案。后文中用于与陶器器形知识本体集成的领域本体模型是笔者基于 CIDOC CRM 本体族构建的考古学领域本体模型，由于陶器器形知识本体中的实体类很少能从现有领域本体模型中复用，因此，该扩展本体不仅能够针对特定领域本体实现交互，且具有一定的可推广性。实现本体集成的关键是在领域本体中选取可交互的实体类和恰当的语义映射方式。本体间的语义映射是异构本体间的桥[26]，常用于本体集成、本体重构及复用和本体模型之间的知识交换[27]。由于类型学本体中的类大多是以关联规则限定的领域本体中已有的实体类，负责其类间语义关联表示的对象属性，也需要复用领域本体模型的关联规则，类内的数据属性主要用于对知识主体添加修饰和约束条件。

类型学的基本逻辑是归纳,从大量的事实中选取包含相应特征的部分归并成类。这种知识逻辑的表达可用到的最基本的语义映射规则是条件限定下的等价映射,在 OWL 中以 EquivalentTo 语句表达,该语句用于表示两个实体类的等价关系,可以通过对象属性或数据属性关系进行条件约束,条件等价映射的基本规则可示意如图 9 所示:

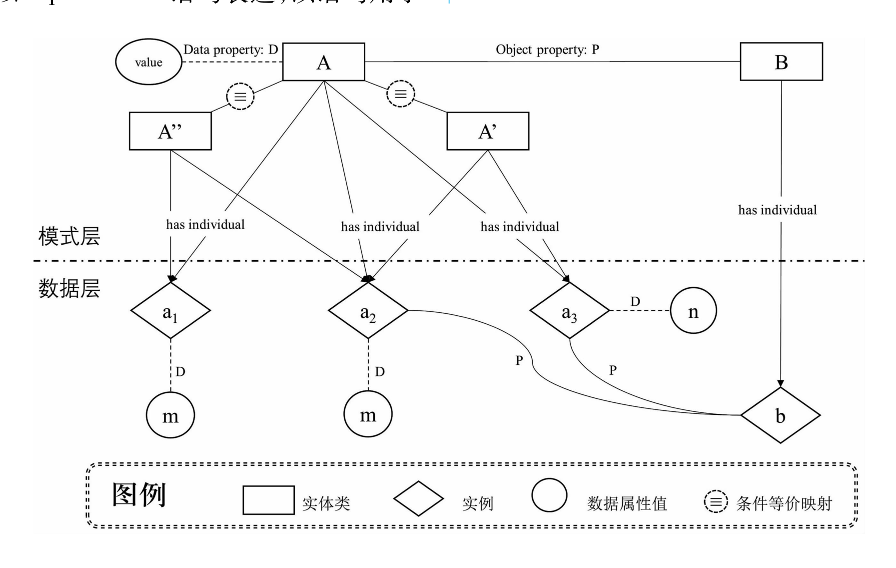


图 9　OWL 条件等价映射的基本规则示意

在图 9 中,有实体类 A,包含实例 a_1、a_2、a_3;和实体类 B,包含实例 b;对象属性 P 的 Domain 为类 A,Range 为类 B;数据属性 D 的领域为类 A,取值为 m、n。

若要将类 A 中所有与实例 b 以对象属性 P 关联的实例提取归纳为类 A′,其映射语句为:

A′ EquivalentTo A and (P value b)

经过本体推理功能,实体类 A’会被推理为类 A 的子类,且包含实例 a_2、a_3。

又若要将类 A 中所有数据属性 D 的属性值为 m 的实例提取归纳为类 A″,其映射语句为:

A′ EquivalentTo A and (D value “m”)

经过本体推理功能,实体类 A″会被推理为类 A 的子类,且包含实例 a_1、a_2。

上述所列的是条件等价映射最基本的形式,在应用中可以定义更为复杂的映射条件进行多层次的嵌套表达,但为了保障规则的稳定性,语义映射的约束条件尽量不超过 3 条。

此外,在前文对陶器类型表示的过程中,设置了某种类型陶器不可能包含的部分 $P_{\overline{V_x}}$,这一设置本身并非单纯的知识表示需要,而是在对陶器类建立本体模型时,以语义关联规则设置排除条件,以有助于规避错误的本体推理结果,提高本体质量,并能在后续的知识计算环节发挥一定的作用。按照公式 3 的规则,在类型学本体模型中对 $P_{\overline{V_x}}$ 限定的方式是对类别 V_x 下的实例添加 Negative object property assertions 条目的方式实现。以图 7 所示的陶器类“鼎”不可能包含默认部分“空足”,可表示为:

not (鼎 P9_consists_of 空足)

接下来,仍以上文中出土陶器的器形表示问题为例,展示考古类型学本体扩展的具体方法,并对语义映射规则进行探讨。

6.2　考古学类型学本体扩展研究

根据前文提出的出土陶器器形的知识表示模型,和类型学本体扩展的语义映射规则,此时的本体扩展任务就是对上述表达式中的各个部分分别进行形式化转换,将其按照语义关联规则与领域本体集成,进而实现类型学知识向领域知识图谱的融合,经过分析,对陶器器形知识表示模型的形式化策略见表 2。

如表 2 所示,按照出土陶器器形的表达式对其中的知识元素逐个定义,扩展模型中新增了两个实体类,分别为表示器形部分 HE92_Object_Component_Part 和表示部分结构位置的 HE93_Object_Component_Position,其他均为以条件等价映射规则限定的,领域本体中原有的实体类,用于对陶器器形部分和陶器类型进行层次表示。以此来实现对考古类型学的形式化语义表示,使其可以与考古学或文物与文化遗产领域的知识图谱进行知识融合。由于数据完整性和篇幅问题,

表 2 出土陶器器型分类知识的形式化表示

部分		$P_{mn}=\{R_L,R_{P_m},F_D,F_R,F_C\}$		类型	说明
		P_mn	arc:HE92_Object_Component_Part	类	陶器部分类
		P_L	arc:HE93_Object_Component_Position	类	部分的结构位置
		P_H	arc:HE931_PH	类	结构位置类:上、中、下、附件
		P_M	arc:HE932_PM	类	· **EquivalentTo** HE92_Object_Component_Part
		P_B	arc:HE933_PB	类	and (RL **value** "H/M/B/A")
		P_A	arc:HE931_PA	类	
		P_{Li}	arc:HE92i_PLi L∈(H,M,B,A),i=1,2,3,…	类	默认部分类 · 示例:"口" HE921_PH1 HE921_PH1 rdfs:comment "口-默认部分-上侧" Class: HE921_PH1 HE921_PH1 SubClassOf HE931_PH HE921_PH1 EquivalentTo HE92_Object_Component_Part and (RPm value "口")
		R_L	RL	数据属性	结构位置标记 · "H","M","B","A"(rdfs:Literal)
		R_{P_m}	RPm	数据属性	默认部分标记 · "口","底","腹"…(rdfs:Literal)
		F_D	FD	数据属性	描述修饰属性 · "方","圜","平"…(rdfs:Literal)
		F_R	FR	数据属性	相对修饰属性 · "大","宽","侈"…(rdfs:Literal)
		F_C	FC	数据属性	定量修饰属性 · 自然数(xsd:integer)
类	类别	$V_x=\{R_{Usage},P_{V_x},P_{\overline{V_x}}\}$			
		V_x	arc: HE911x_Vx	类	类别类 · 示例:"壶" HE9116_V6 HE9116_V6 SubClassOf HE911_Pottery_Product Class: HE9116_V6 HE9116_V6 EquivalentTo HE9_Typological and (RVx value "壶") HE9116_V6 rdfs:comment "壶-陶器器型"
		R_{Usage}	arc: HE96_Usage	类	功用类
		P_{V_x}	对象属性筛选		必须包含的部分 · P9_forms_part_of some HE92i_PLi
		$P_{\overline{V_x}}$			不可能包含的部分 · **not** (P9_forms_part_of some HE92i_PLi)
型	型别	$V_{xy}=\{R_{V_x},P_{V_{mn}}\}$			
		V_{xy}	arc: HE911xy_Vxy	类	型别类 · 示例:"腹耳壶" "腹耳壶" HE91164_V64 HE91164_V64 EquivalentTo HE9116_V6 and (P9_consists_of value 耳-2) Class: HE91164_V64 HE91164_V64 rdfs:comment "腹耳壶" HE91164_V64 SubClassOf HE9116_V6
		R_{V_x}	RVx	数据属性	类别标记 · "豆","鼎","壶"…(rdfs:Literal)
		$P_{V_{mn}}$	对象属性筛选		包含型别特征部分 · P9_forms_part_of some HE92_Object_Component_Part
	式别	$V_{xyi}=\{R_{V_{xy}},P_{V_{xyi}},O_{V_{xyi}},S_{V_{xyi}}\}$			本文对陶器式别及更细的层次划分不做继续延展,以实例表示
		V_{xyi}	**IndividualOf** arc:HE9_Typological	实例	

表 2 未对式别表达式进行展开,而将其直接作为实例处理,对其进行形式化表示的方式与型别大同小异。

类型学扩展本体与文化遗产领域通用性最强的 CIDOC CRM 模型的交互位置为 HE92_Object_Component_Part 与 E19_Physical_objects 之间的关联,二者之间的直接关联以 P9_forms_part_of(consists_of)表示。HE93_Object_Component_Position 之下,定义了分别用于表示上侧、中间、下侧、附件四种结构位置的子类,分

韩牧哲，高劲松，李钰. 面向考古类型学的出土陶器器形的知识表示与语义关联构建[J]. 图书情报工作,2022,66(12):92－107.

别为 HE931、HE932、HE933、HE934，在一般应用中，这四个子类并不直接参与知识表示，但是它们在知识计算过程中可以作为知识结构一致性表达的框架节点，决定知识结构树的分支情况；另外在实际应用中，陶器的部分的结构位置可以帮助对一些出土陶器的残缺部分进行知识表示。

7 出土陶器器形知识表示的实证

本文仍以图 8 所示青海柳湾的两件陶器为例进行陶器器形分类知识表示的效果，两件陶器的线图和特征部分已在图中展现。示例的两件彩陶腹耳壶均取自青海柳湾墓地，文物发掘编号分别为柳湾 M528:1 和柳湾 M578:4，均属于马家窑文化，文化分期分别为半山类型早期和马厂类型早期。为方便表示，本文对原报告的类型划分进行简化，简化后其分类结构如图 10 所示：

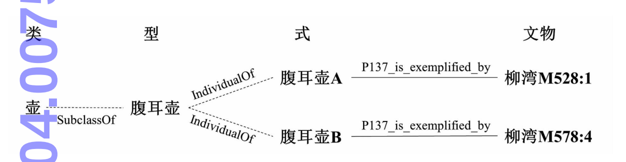


图 10 示例柳湾腹耳壶类型划分示意

图 10 中，陶器类别为“壶”（V_6），陶器中的壶，是一种非常普遍的盛贮器，其必须包含的部分为“口”“颈”“腹”“底”，且一般不限定其不可包含的部分。因此，对实例两件文物的类别“壶”的知识表示为：

$$V_6 = \begin{pmatrix} R_{Usage} & 盛贮器 & & & \\ P_{(V_x} & 口 & 颈 & 腹 & 底 \\ P_{\overline{V_x}} & \emptyset & & & \end{pmatrix}$$

型别为“腹耳壶”（V_{64}），其关联的上位类为“壶”（V_6），其型别特征为“双腹耳”（P_{A11}），可表示为：

$$V_{64} = \begin{pmatrix} R_{V_x} & 壶 \\ P_{V_{mn}} & 双腹耳 \end{pmatrix}$$

式别分别为“腹耳壶 A”（V_{64A}）和“腹耳壶 B”（V_{64B}），省略轮廓信息和规格信息，其表达式分别为：

$$V_{64A} = \begin{pmatrix} R_{V_{xy}} & 腹耳壶 & & & & \\ P_{V_{xyi}} & 直口 & 直颈 & 圆肩 & 扁鼓腹 & 双腹耳 & 平底 \end{pmatrix}$$

$$V_{64B} = \begin{pmatrix} R_{V_{xy}} & 腹耳壶 & & & & \\ P_{V_{xyi}} & 侈口 & 短颈 & 广肩 & 深腹 & 双腹耳 & 平底 \end{pmatrix}$$

对于 V_{64A} 和 V_{64B} 中的出现的具体部分而言，以“短颈”（P_{H21}）、“扁鼓腹”（P_{M221}）、“双腹耳”（P_{A11}）三例加以说明：

$$P_{H21} = \begin{pmatrix} R_L & H \\ R_{P_m} & 颈 \\ F_D & \emptyset \\ F_R & 短 \\ F_C & 1 \end{pmatrix}; P_{M221} = \begin{pmatrix} R_L & M & \\ R_{P_m} & 腹 & \\ F_D & 扁 & 鼓 \\ F_R & \emptyset & \\ F_C & 1 & \end{pmatrix};$$

$$P_{A11} = \begin{pmatrix} R_L & A \\ R_{P_m} & 耳 \\ F_D & 腹耳 \\ F_R & \emptyset \\ F_C & 2 \end{pmatrix} \cdots\cdots$$

将分层次的特征向量按照表 2 所示的关联规则转化为语义信息，示例中的两件陶器即可形式化表示见表 3。

将上述数据录入本体编辑工具 Protégé5.5.0，使用 OntoGraf 工具生成的本体图形界面如图 11 所示。图中按照 CIDOC CRM 的本体概念层次对实体类进行了标识，如对陶器器形展开到“型别”（HE91164_V64，腹耳壶）层面，则扩展后的领域本体模型的层次可延展至 8 层。

8 结语

从类型学视角出发对出土文物进行精确描述和量化比对的思维自现代考古学发轫之时已被提出，但直至今日仍缺乏有效的实际方法。本文以考古类型学的固有逻辑为指导，尝试对出土陶器的器形从属种关系、整部关系两种维度进行描述，并运用特征向量模型对陶器的不同部分和不同类型层次进行知识表示，进而构建了面向类型学的出土陶器器形知识表示模型。在此基础上，提出了考古类型学本体扩展方法，以条件等价映射方法实现扩展本体与现存领域本体的集成，促进考古类型学知识与领域知识的融合和共享。以陶器器形为例对出土文物从考古类型学视角下的知识表示是数字人文研究在考古学领域的一种新的尝试，有望在应用中为考古工作者提供一致的结构化术语词表和形式化的类型学本体模型，相关方法有助于高效解决出土文物类型深层次、细粒度表示问题，在此基础上的知识应用和知识服务有望促进考古学领域的知识共享并提升重用效果。数字人文在考古学研究领域中的应用尚处于萌芽状态，当前的研究尚有很多可进一步探究的方向，由于缺乏较大规模的公开数字语料集，本文研究过程中仅针对性地采集并组织了少量数据进行实验，后续在知识组织方面有必要在拥有成熟数据集的基础上，尝试解决对考古类型学语料的大规模自动语

表 3　示例柳湾腹耳壶的形式化描述效果列表

柳湾 M528:1			
P137_exemplifies	**柳湾 – 腹耳壶 A**	RVxy “腹耳壶”，RVx “壶”	
	P9_consists_of	**口 – 2**	rdfs: comment “直口”，RL “H”，RPm “口”，FD “直”，FC 1
	P9_consists_of	**颈 – 1**	rdfs: comment “直颈”，RL “H”，RPm “颈”，FD “直”，FC 1
	P9_consists_of	**肩 – 3**	rdfs: comment “圆肩”，RL “H”，RPm “肩”，FD “圆”，FC 1
	P9_consists_of	**腹 – 12**	rdfs: comment “扁鼓腹”，RL “M”，RPm “腹”，FD “扁”，FD “鼓”，FC 1
	P9_consists_of	**底 – 1**	rdfs: comment “平底”，RL “B”，RPm “底”，FD “平”，FC 1
	P9_consists_of	**耳 – 2**	rdfs: comment “双腹耳”，RL “A”，RPm “耳”，FD “腹耳”，FC 2
柳湾 M578:4			
P137_exemplifies	**柳湾 – 腹耳壶 B**	RVxy “腹耳壶”，RVx “壶”	
	P9_consists_of	**口 – 1**	rdfs: comment “侈口”，RL “H”，RPm “口”，FR “侈”，FC 1
	P9_consists_of	**颈 – 3**	rdfs: comment “短颈”，RL “H”，RPm “颈”，FR “短”，FC 1
	P9_consists_of	**肩 – 1**	rdfs: comment “广肩”，RL “H”，RPm “肩”，FD “广”，FC 1
	P9_consists_of	**腹 – 3**	rdfs: comment “深腹”，RL “M”，RPm “腹”，FR “深”，FC 1
	P9_consists_of	**底 – 1**	rdfs: comment “平底”，RL “B”，RPm “底”，FD “平”，FC 1
	P9_consists_of	**耳 – 2**	rdfs: comment “双腹耳”，RL “A”，RPm “耳”，FD “腹耳”，FC 2

注：表中粗体内容代表实例；斜体内容代表对象属性，其上方实例为 Domain，右侧实例为 Range；正常字体为数据属性及其属性值，其左侧实例为 Domain

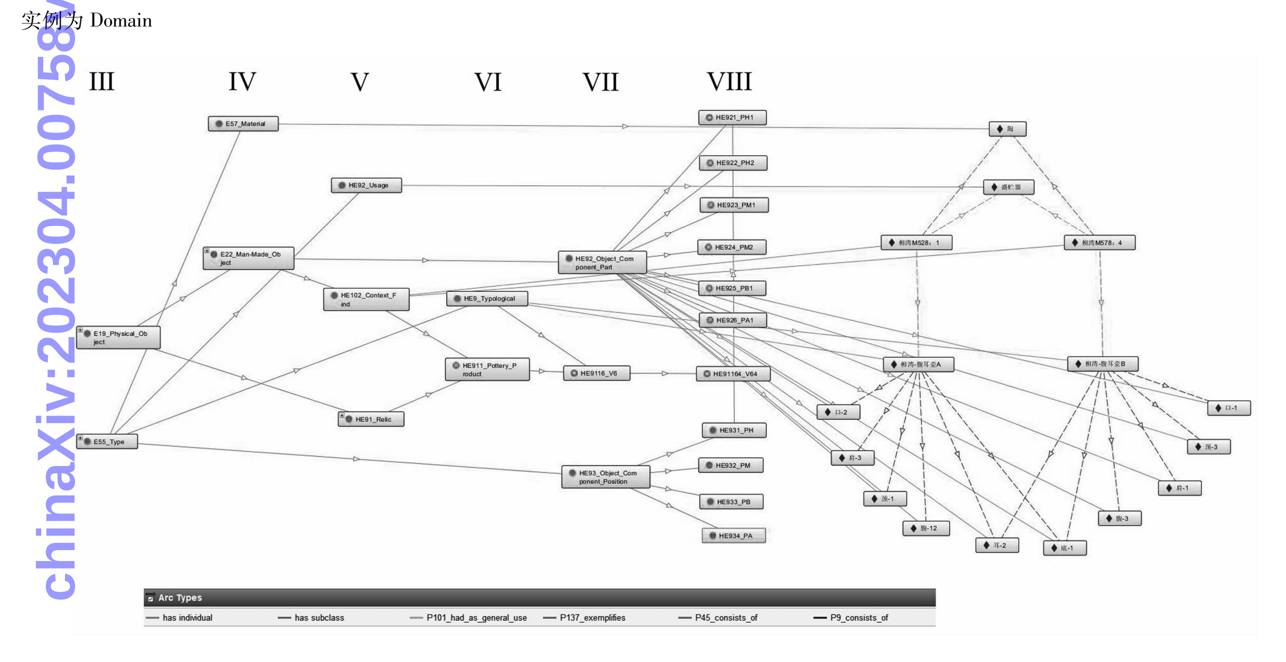


图 11　陶器器形分类知识本体图形可视化

义标注和知识抽取问题；在知识服务方面，需要探讨基于语义相似度计算的考古类型学应用的数字实现问题。

参考文献：

[1] 蔡迎春. 数字人文评价：学科性、专业性、技术性[J]. 中国图书馆学报，2021，47(4)：86 – 96.

[2] 张昱杰. 考古数据信息提取和分析技术的研究与应用[D]. 杭州：浙江大学，2019.

[3] 高劲松，韩牧哲. 考古发掘资料图数据库的语义关联构建研究[J]. 图书情报工作，2021，65(9)：105 – 116.

[4] 何庆. 基于 GIS 的考古勘探信息系统研究与应用[J]. 测绘与空间地理信息，2019，42(12)：143 – 145.

[5] 俞伟超. 考古类型学的理论与实践[M]. 北京：文物出版社，1989.

[6] ALMEIDA B，ROCHE C，COSTA R. Archaeological classification and ontoterminology：the case of Islamic archaeology of the al-Andalus[C]// TOTh 2017. Terminologie & Ontologie：Théories et Applications，Chambéry，France. 2017：221 – 236.

[7] CIDOC-CRM[DB/OL]. [2021 – 09 – 14]. http://www.cidoc-crm.org/.

[8] CDWA[DB/OL]. [2021 – 09 – 14.] http://www.getty.edu/research/publications/electronic_publications/cdwa/.

[9] VRA Core[DB/OL]. [2021 – 09 – 14]. http://www.loc.gov/standards/vracore/.

[10] 文物数字化保护元数据标准规范征求意见稿发布[EB/OL]. [2021 – 09 – 14]. https://www.lib.pku.edu.cn/portal/cn/news/0000001494.

[11] 张开兴，王丽，刘贤喜，等. 基于本体的水稻收割机底盘虚拟装配工艺知识库设计[J]. 农机化研究，2015(11)：121 – 124.

[12] 王莹，戴毅茹，王坚. 基于三维框架本体建模的钢铁生产工艺知识库系统设计与实现[J]. 电脑知识与技术，2017(2)：242－244.
[13] 李长杰，明新国，邱坤华，等. 基于本体的飞机工装设计知识表示方法[J]. 中国机械工程，2014，25(19):2614－2619.
[14] 任民山，蔡红霞，李钧. 基于本体的民机装配工艺知识库系统构建[J]. 工业控制计算机，2018，31(4)：122－123.
[15] 丘宏俊，陶华，高晓兵，等. 飞机装配工艺设计知识本体[J]. 西北工业大学学报，2007，25(1):32－36.
[16] 张发平，李丽. 基于多维层次情境模型的业务过程知识推送方法研究[J]. 计算机辅助设计与图形学学报，2017，29(4):751－758.
[17] 张光直. 考古学专题六讲[M]. 北京：生活・读书・新知三联书店，2013：57.
[18] 崔帅. 非经典逻辑视阈下的知识表征分析[D]. 太原:山西大学,2018.
[19] 新疆文物考古研究所. 新疆察吾呼——大型氏族墓地发掘报告[M]. 北京:东方出版社，1999.
[20] 中国社会科学院考古研究所. 中国考古学・新石器时代卷[M]. 北京：中国社会科学出版社，2010.
[21] 梁思永. 山西西阴村史前遗址的新石器时代的陶器[J]. 三晋考古，1996(00):300－338,351－354.
[22] 李济. 西阴村史前的遗存[M]. 北平:清华学校研究院，1927.
[23] 青海省文物管理处. 青海柳湾[M]. 北京:文物出版社，1984.
[24] 周明全，樊亚春，耿国华. 一种基于空间对称变换的三维模型形状描述方法[J]. 电子学报，2010，38(4):853－859.
[25] 钱建军. 基于语义的唐卡图像标注与检索研究[D]. 兰州:西北民族大学. 2010.
[26] 王昊奋，漆桂林，陈华钧. 知识图谱:方法、实践与应用[M]. 北京：电子工业出版社，2019：193－195.
[27] 张凌宇，陈淑鑫，张新. 一种基于上下文的语义映射方法[J]. 计算机应用研究，2014，31(10):2990－2993,2997.

作者贡献说明：

韩牧哲:提出研究思路,设计研究方案,实验与分析,论文撰写;
高劲松:设计研究方案,论文指导,论文修改;
李钰:数据采集与整理,参与实验,参与论文撰写与修改。

Knowledge Representation and Semantic Association Construction of the Shape of Unearthed Potteries for Archaeological Typology

Han Muzhe[1] Gao Jinsong[1] Li Yu[2,3]
[1] School of Information Management, Central China Normal University, Wuhan 430079
[2] School of Archaeology and Museology, Sichuan University, Chengdu 610225
[3] Academy of Art and Design, Sichuan University of Culture and Arts, Mianyang 621000

Abstract: [Purpose/Significance] On the view of archaeological typology, this paper proposes a knowledge representation model suitable for the description of the characteristics of unearthed relics, and the corresponding semantic mapping and ontology expansion methods are proposed on this basis, which may break through the semantic barriers caused by traditional typology methods and promote the sharing and reuse of knowledge. [Method/Process] Firstly, this paper analyzed the semantic barrier caused by the thought and traditional methods of archaeological typology, and presented a targeted archaeological typology knowledge representation strategy; Secondly, combined with the characteristics of archaeological corpus and the logic of typology, and the shape description of unearthed potteries was decomposed according to two dimensions of genus-species relationship and whole-part relationship; Next, the knowledge representation scheme and eigenvector expressions of the parts and varieties of potteries were put forward layer by layer, and then put forward the knowledge representation model of the shape of unearthed potteries; And, based on the knowledge representation model, an ontology extension method of archaeological typology ontology based on conditional equivalence mapping was proposed to realize the construction of semantic association of archaeological typology; Finally, taking two samples from Liuwan Cemetery, Qinghai Province as example, the paper showed the form and process of typological knowledge representation and the visualization effect of ontology graphics. [Result/Conclusion] Taking the shape of potteries as a case, the knowledge representation and semantic association construction of unearthed relics from the perspective of archaeological typology is a new attempt of digital humanities research in the field of archaeology, which can provide technical support for the transformation of typological research from relying on experience to relying on data.

Keywords: archaeological typology　unearthed potteries　knowledge representation　semantic association construction